FISHES OF TEXAS PROJECT: DATA VISUALIZATION AND ANALYSIS TOOLS

Colton Avila, Dean A. Hendrickson, Adam E. Cohen, Melissa Casarez

University of Texas at Austin, Integrative Biology, Biodiversity Center (TNHC)

ABSTRACT

The Fishes of Texas Project (FoTX) (http://fishesoftexas.org) database currently has 124,452 specimen-vouchered occurrence records spanning > 150 years with over 400,000 new records (including non-vouchered sources such as literature, anecdotal, and photo-based) in the process of being imported.

Continual data growth prompted creation of new tools to dynamically assess (as the data evolve) the state of data coverage across various dimensions to increase user understanding and accessibility to the data and improve overall utility of the project.

We produced species sampling curves, temporal species accumulation graphs, and heat maps of collecting event density over time and space for each river sub-basin within Texas. A QGIS plugin was also created to better assess the suspect status of incoming records.

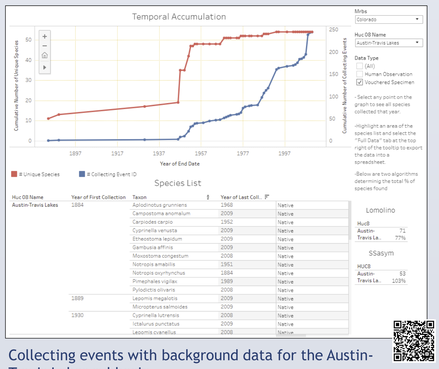
Each type of visualization has basic documentation, easily accessible statistical summaries, flexible queries, and exploration tools to help reveal variations in sampling density over both temporal and spatial dimensions.

We highlight here the San Bernard River as an example of a notably under-sampled sub-basin (as indicated by diverse forms of evidence).

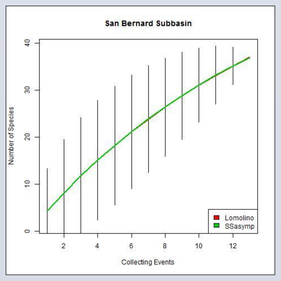
With addition of future records, these dynamic tools will continue to illustrate taxonomic and spatial sampling deficiencies that in turn will help guide conservation planning.

SPECIES ACCUMULATION

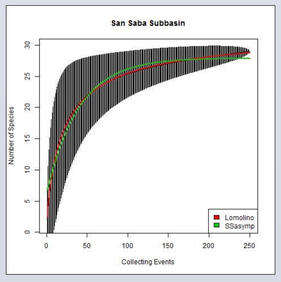
- Allows the user to discover periods of time when sampling effort was maximized
- Pinpoints collecting events that produced species not previously recorded
- Generates an asymptote when nearing the maximum species richness for a sub-basin



Collecting events with background data for the Austin-Travis Lakes subbasin



Sub-basin estimated to see a large increase in species richness with increased sampling effort



Sub-basin estimated to have hit species richness asymptote

Import and clean data

↓

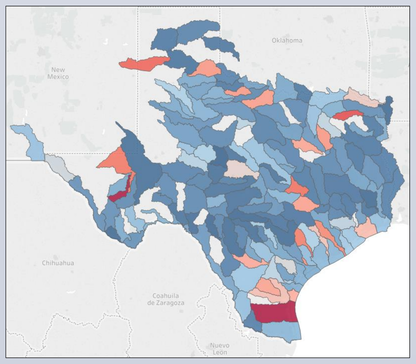
Iterate over sub-basins and species found within

↓

Species accumulation models from vegan package in R (Lomolino / Ssasymp)

↓

Shade sub-basins by sampling completeness



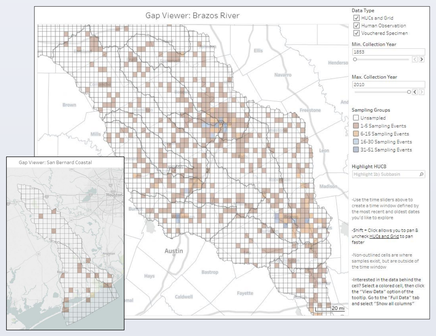
Sub-basins shaded by estimated sampling completeness

- Highlights sub-basin most in need of sampling
- Red indicates sub-basins that have the highest probability for new species to be found
- Blue indicates sub-basins nearing the asymptote for species richness
- Empty sub-basins indicate the calculation failed due to under-sampling or insufficient species diversity

HEAT MAPS

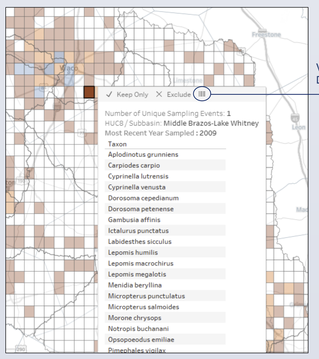
- 8-Digit Hydrologic Units (HUC8s) for all Texas major river basins were taken from the USGS Watershed Boundary Dataset, HUC8s were used to delineate sub-basins within the greater basin
- Each HUC8 was broken into a grid of 3 mile square polygons
- The underlying record data is connected to the FoTX database and sub-queried into extracts for faster visualization speed
- Before overlaying FoTX records with the shapefiles, points with large location uncertainty (greater than 3 miles), or flagged as "suspect" (as defined by FoTX), were omitted

Gap Viewer



Sampling gap heatmaps: the San Bernard Coastal gap viewer (left) illustrates an overall under sampled sub-basin, while the Brazos River gap viewer (right) highlights varying concentrations of sampling throughout the basin

- Users can filter data based on data type, collection year, and number of records by area
- Highlighting sub-basins within the major drainage basin helps break apart clustered collection areas
- Under-sampled areas are easily identifiable



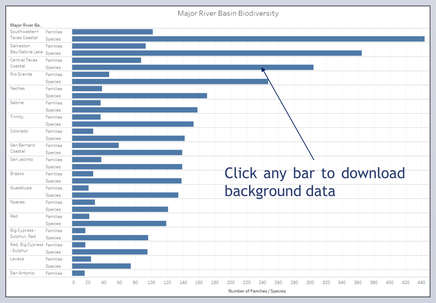
Tooltip for a cell SE of Waco

View Background Data

Records and associated information that make up a cell

- The ability to highlight an area of interest allows the user to spatially query the data
- The background data for any cell can be downloaded
- These visualizations are quick to search through, with a streamlined data update process for new occurrence records
- Increases data accessibility - novice users can now explore the data without requiring the geographic or taxonomic knowledge used in text-only search methods
- Future efforts look to improve spatial scaling and maintain continuous database connectivity instead of using extracts

DATA SUMMARY

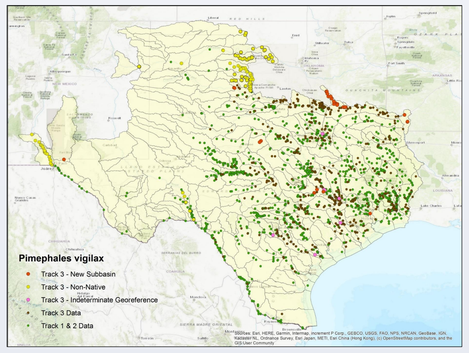


Measuring family and species diversity within the major river basins in Texas

- Summarizes:
  - Sub-basin biodiversity
  - Record quality
  - Collecting events
  - Collectors
  - Data type
- Allows users to download and manipulate data behind the view

DATA QUALITY EVALUATOR

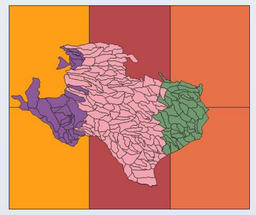
This QGIS plugin compares incoming occurrence data to existing FoTX records for each HUC8. We developed this tool to automate the process of finding and flagging "suspect" records. Incoming records are "suspect" if any of the following apply:



Historic vouchers of Pimephales vigilax : 345 out of 1294 records were identified as suspect

1. It is the first occurrence of the species in the HUC8
2. It falls outside of Fishes of Texas' determination for the native range of the species
3. It's coordinate uncertainty could place it in neighboring sub-basins where the species has yet to be found

All records were projected into the UTM coordinate system in order to be buffered by their locality uncertainty (in meters for Darwin Core compliance). This was done to better compare the points to their respective HUC8s. Records with an uncertainty larger than 28 kilometers were omitted.



Texas needed to be broken into three different Universal Transverse Mercator (UTM) zones

- In many cases this tool reduced the number of records in need of manual assessment by over 75%. One set of 44,756 vouchered records was reduced to 17,290, of which 9,034 were flagged as suspect due to non-nativity determination
- It also highlights occurrence records that might be overlooked due to proximity to heavily collected areas
- The locality suspect flag would be applied only if the record, after being buffered, had less than 25% of its buffered area located within the record's listed HUC8
- This tool increases both the rate at which data can be processed and published and the quality of suspect determinations for FoTX records, which facilitates discussion regarding the extension of species ranges based on non-native determination

SUMMARY/DISCUSSION

- Visually highlights basins that should be prioritized in terms of conservation efforts
- Facilitates conservation planning discussion by enabling inclusion of those with limited geographic and taxonomic knowledge
- Generally increases accessibility to, and utilization of, FoTX data in research and management

ACKNOWLEDGEMENTS / CONTACT

A collaborative project of the University of Texas at Austin and Texas Parks and Wildlife Department, under the U.S. Fish and Wildlife Service State Wildlife Grant Program, grant TX T-180-R-1, F17AF01129 (CFDA# 15.634)

Special thanks to Tomislav Urban (Texas Advanced Computing Center – TACC)

fishesoftexas@gmail.org